

Serial No.: 09/929,094

### AMENDMENT TO THE CLAIMS

1. (Previously Presented) The method according to Claim 28, further comprising:

constructing a user cluster index for the user;

wherein the user cluster index comprises a list of families of data to which data from the digital data gathering results of the user were categorized;

monitoring families of the further digital data gathering results of the user; and

comparing the families of the further digital data gathering results of the user to the user cluster index to determine anomalies in the digital data gathering results.

2. (Previously Presented) The method according to Claim 1, further comprising:

comparing the anomalies to the user cluster index to determine the ratio of anomalies to existing clusters; and

reporting a potential misuse when the ratio exceeds a predetermined threshold.

Serial No.: 09/929,094

3. (Previously Presented) The method according to Claim 1, further comprising:

constructing a user lexicon for the user;

wherein the user lexicon comprises a list of words or phrases gathered from the digital data gathering results of the user; and

comparing words or phrases gathered from the further digital data gathering results to the user lexicon to determine anomalies in the digital data gathering results.

4. (Previously Presented) The method according to Claim 3, wherein the user lexicon further comprises a list of words or phrases gathered from the monitoring of the queries; and

comparing the further content of the further query to the user lexicon to determine any anomaly by the further query.

5. (Previously Presented) The method according to Claim 3, further comprising:

determining a ratio of anomalies to words or phrases in the lexicon; and

reporting a potential misuse when the ratio exceeds a predetermined threshold.

6. (Previously Presented) The method according to Claim 3, wherein the user lexicon comprises a list of words or word strings identifying particular words or types of words, or both, extracted from documents returned in response to user queries.

Serial No.: 09/929,094

7. (Previously Presented) The method according to Claim 1, further comprising:

constructing a structured data profile for the user;

wherein the structured data profile comprises a list of data identifying employment characteristics of the user;

comparing the further digital data gathering results of the user to the structured data profile to determine whether the further digital data gathering results are congruent with the structured data profile; and

identifying a potential misuse when the digital data gathering results are not congruent with the structured data profile.

8. (Previously Presented) The method according to Claim 7, wherein the structured data profile comprises a structured data profile lexicon of terms and phrases indicating valid user activity.

9. (Previously Presented) The method according to Claim 3, further comprising:

constructing a structured data profile for the user;

wherein the structured data profile comprises a list of data identifying employment characteristics of the user;

comparing the further digital data gathering results of the user to the structured data profile to determine whether the further digital data gathering results are congruent with the structured data profile; and

identifying a potential misuse when the further digital data gathering results are not congruent with the structured data profile.

Serial No.: 09/929,094

10. (Currently Amended) A method for identifying the misuse of authorized access to a digital data gathering system by a user, comprising:

monitoring a content of digital data gathering results of the user, wherein the content includes at least one of words and phrases;

constructing a user lexicon for ~~[[a]] the user of [[a]] the digital data gathering system;~~

wherein the user lexicon comprises a list of ~~at least some of the a plurality of~~ words or phrases gathered from documents of the digital data gathering results of the user;

monitoring a further content of further digital data gathering results obtained by the user;

comparing words or phrases of the further content to the user lexicon to determine anomalies in the further digital data gathering results; and

identifying a potential misuse when an anomaly is detected.

11. (Previously Presented) The method according to Claim 10, further comprising:

determining a ratio of anomalies to the words or phrases in the lexicon;

and

reporting a potential misuse when the ratio exceeds a predetermined threshold.

12. (Previously Presented) The method according to Claim 10, wherein the user lexicon comprises a list of words or word strings identifying nouns extracted from documents returned in response to user queries.

Serial No.: 09/929,094

13. (Currently Amended) The method according to Claim 10, further comprising:

constructing a structured data profile for ~~[[a]] the user of [[a]] the~~ digital data gathering system;

wherein the structured data profile comprises a list of data identifying employment characteristics of the user;

comparing the further digital data gathering results of the user to the structured data profile to determine whether the further digital data gathering results ~~are congruent~~ correspond with the structured data profile; and

identifying a potential misuse when the further digital data gathering results ~~are not congruent~~ do not correspond with the structured data profile.

14. (Currently Amended) A method for identifying the misuse of authorized access to a digital data gathering system by a user, comprising:

constructing a structured data profile for ~~[[a]] the user of [[a]] the~~ digital data gathering system;

wherein the structured data profile comprises a list of data identifying employment information of the user;

monitoring digital data gathering results of the user;

comparing digital data gathering results of the user to the structured data profile to determine whether the digital data gathering results ~~are congruent~~ correspond with the structured data profile; and

identifying a potential misuse when the digital data gathering results ~~are not congruent~~ do not correspond with the structured data profile.

Serial No.: 09/929,094

Claims 15-16 (Canceled)

17. (Previously Presented) The method according to Claim 29, further comprising: weighting anomalies identified according to the user lexicon, the user cluster index, and the structured data profile to determine a report of potential misuse.

18. (Currently Amended) The method according to Claim 29, further comprising: sending a notification of potential misuse when [[a]] an anomaly is identified according to two or more of the user lexicon, the user cluster index, and the structured data profile.

19. (Previously Presented) The method according to Claim 29, wherein the user lexicon comprises a list of words or phrases gathered from metadata of documents returned in the query results.

20. (Previously Presented) The method according to Claim 29, wherein the user lexicon comprises a list of words, or types of words, or both, extracted from documents returned in the query results.

21. (Previously Presented) The method according to Claim 29, wherein the user cluster index comprises a list of families of topic data to which the data of the user information retrieval results have been categorized.

Serial No.: 09/929,094

22. (Currently Amended) A method for detecting misuse by a user of an information retrieval system having a document collection, wherein documents of the document collection are categorized into one of a plurality of clusters according to topic, the method comprising:

tracking the one of the plurality of clusters from which any document read by the user originates;

building up a profile of use for the user based on most frequently accessed clusters ~~over a time sufficient to establish a confidence threshold for validity of the profile of the user;~~

tracking each time the user retrieves and reads a document outside of the most frequently accessed clusters; and

establishing a misuse threshold number for documents read outside of the most frequently accessed clusters and after the misuse threshold number is obtained, signaling that a potential misuse may have occurred.

23. (Currently Amended) A method for detecting misuse by a user of an information retrieval system having a document collection, comprising the steps of:

retrieving documents in response to user queries;

clustering the retrieved documents into clusters by category based upon a content of each of the retrieved documents, wherein the content includes at least one of terms, phrases, and topics;

establishing and obtaining a threshold number of retrieved documents and after the threshold number of retrieved documents is obtained, determining a size

Serial No.: 09/929,094

for each of the clusters, and further denoting clusters ~~of a large enough~~ having at least a predetermined size as valid clusters; and

determining if a sufficient predetermined number of retrieved documents do not participate in any of the valid cluster clusters and if not, signaling that a potential misuse may have occurred.

24. (Currently Amended) A method for detecting misuse by a user of an information retrieval system having a document collection, comprising the steps of:

identifying top weighted terms from documents retrieved by the user from searches of the document collection and storing the top weighted terms in a user-specific lexicon;

tracking user activity until the rate of new terms added slows and the user-specific lexicon stabilizes to form a user profile;

identifying for each new query, if the top weighted terms are in the user-specific lexicon;

tracking a ratio of newly occurring terms to terms existing in the user-specific lexicon terms; and

if the ratio of newly occurring terms to existing user-specific lexicon terms exceeds a threshold, signaling that a potential misuse may have occurred.



Serial No.: 09/929,094

25. (Previously Presented) The method according to Claim 24, further comprising:

- tagging the documents to identify words in the documents by type;
- running an original query of terms and phrases;
- selecting specific types of words from relevant documents retrieved by the original query and adding these terms to a second query; and
- iteratively selecting specific types of words from relevant documents retrieved by each query and adding the selected specific types of words to a further query to filter the user-specific lexicon.

26. (Currently Amended) A method for detecting misuse by a user of an information retrieval system having a document collection, comprising the steps of:

- identifying structured data sources that ~~can be~~ are used to identify what the user is working on;
- querying these sources and, for each source, mapping a structured result into a structured data lexicon of terms and phrases that indicate valid user activity;
- for each new query, tracking a ratio of terms found in the structured data lexicon to those not found in the structured data lexicon; and
- if the ratio exceeds a threshold, signaling that a misuse may have occurred.

Serial No.: 09/929,094

27. (Currently Amended) A method for detecting misuse by a user of an information retrieval system having a document collection, comprising the steps of:

identifying structured data sources that ~~can be~~ are used to identify what the user is working on;

querying the identified structured data sources and, for each source queried, mapping a structured result into a structured data lexicon of terms and phrases that indicate valid user activity;

for each new query, retrieving relevant documents for that new query;

extracting key terms from the relevant documents;

identifying the ratio of key retrieved terms found in the lexicon to those not found in the lexicon; and

if the ratio exceeds a threshold, signaling that a misuse may have occurred.

Serial No.: 09/929,094

28. (Currently Amended) A method for identifying a misuse of an authorized user of an information retrieval system, the method comprising:

monitoring a content of a plurality of at least one of queries entered by the user and digital data gathering results obtained by the user, wherein the content includes at least one of terms, phrases, and topics;

constructing a profile of use for the user using the content;

monitoring a further content of at least one of a further query entered by the user and further digital data gathering results obtained by the user;

comparing the further content to the profile of use to determine whether the at least one of the further query entered by the user and the further digital data gathering results is an anomaly;

identifying a potential misuse when an anomaly is detected.

29. (Previously Presented) The method according to Claim 28, wherein the profile of use comprises a user lexicon of user result words or phrases, a user cluster index of result document topic categories, and a structured data profile of known user characteristics, and further comprising:

comparing the further content to each of the user lexicon of user result words or phrases, the user cluster index, and the structured data profile.